

# Ultra boosted HH production study using ML

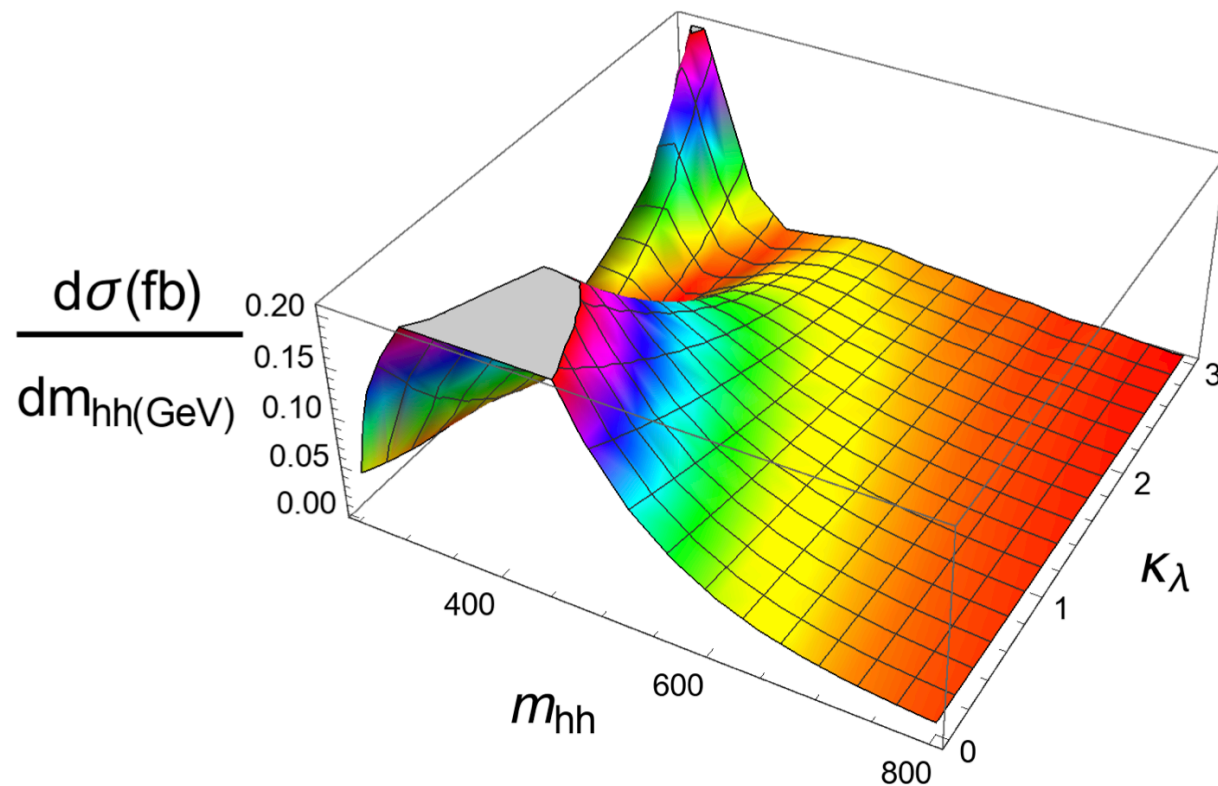
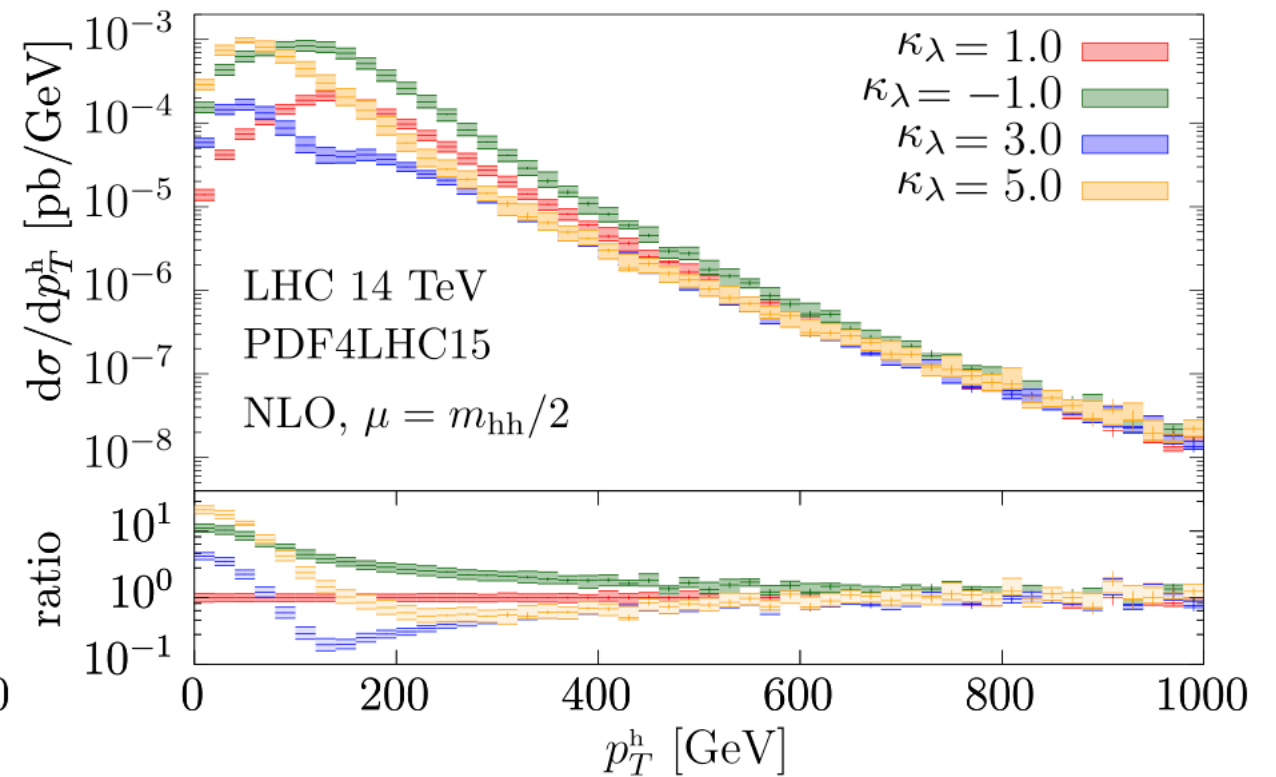
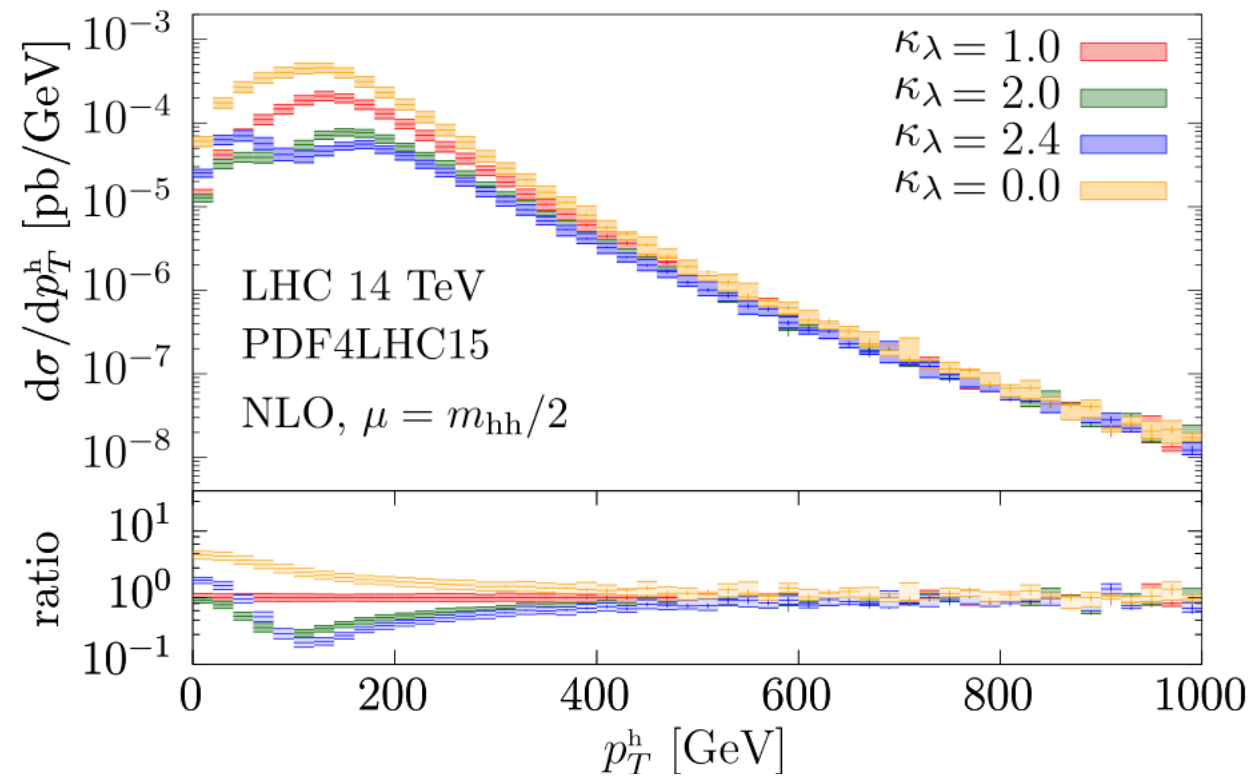
Daniel Diaz, Javier Duarte, Sanmay Ganguly,  
Samadrita Mukherjee, Brian Sheldon

01/09/2021



UC San Diego

# Di-Higgs production at higher energies

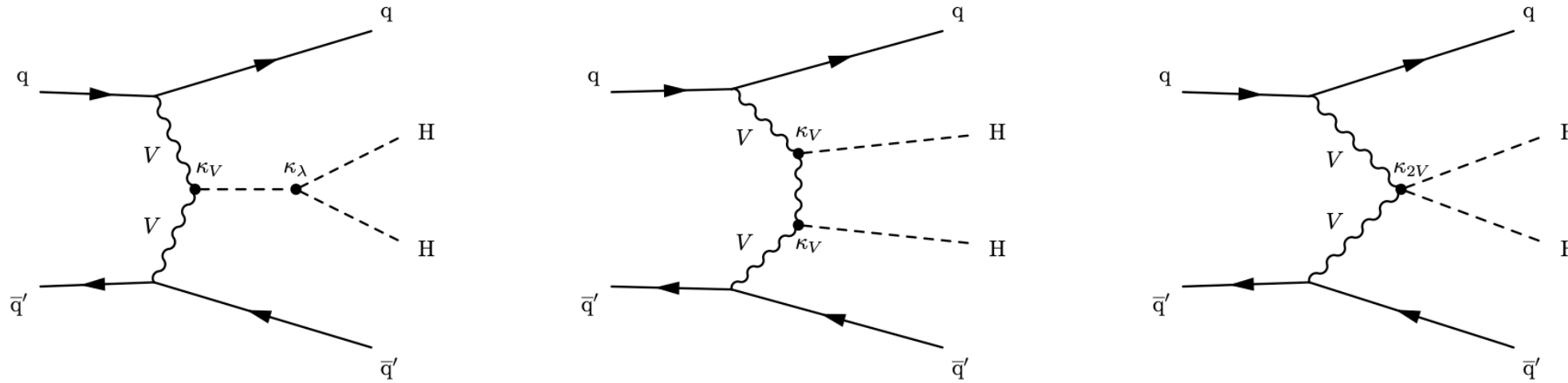


arXiv : 1910.00012

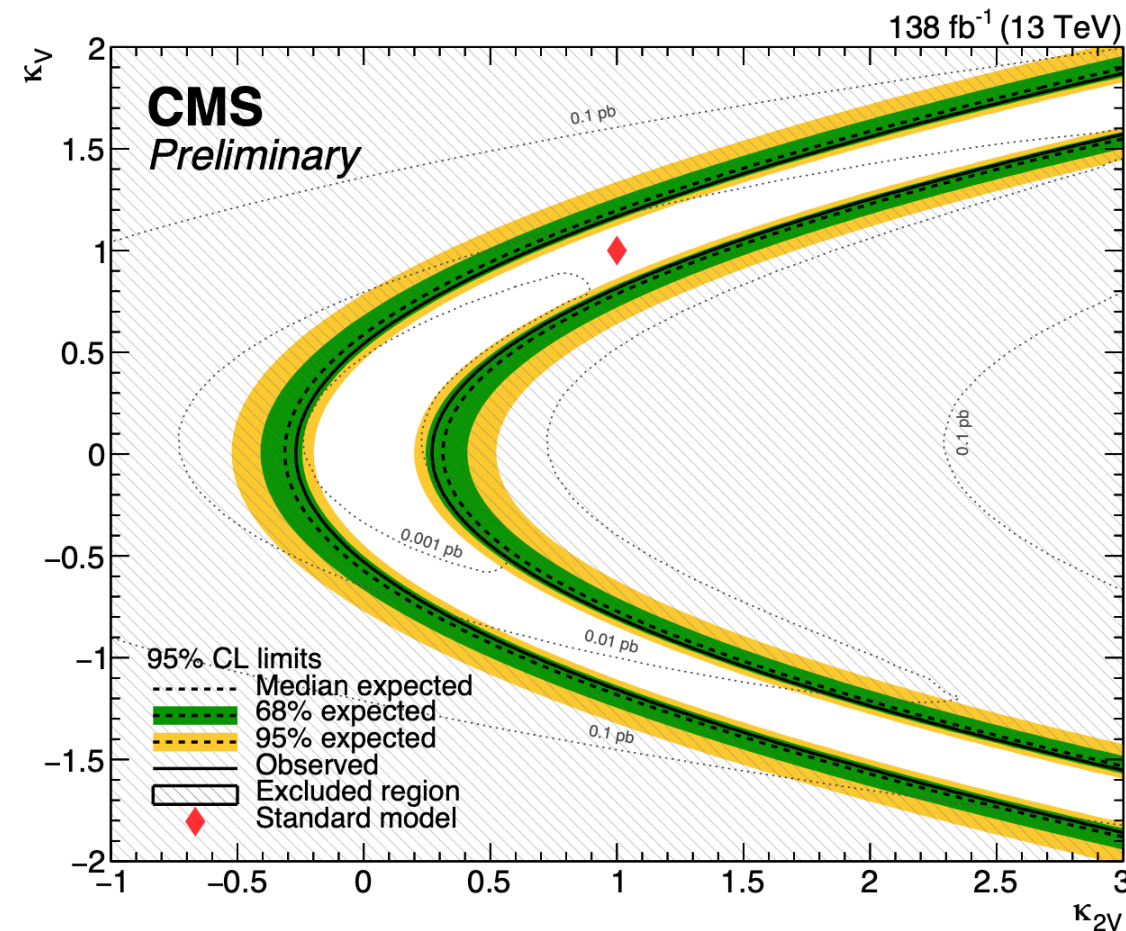
Major backgrounds are t-tbar, QCD

# Recent ML application in HH study

$H \rightarrow b\bar{b}$  tagging is done using ParticleNet (GNN network)

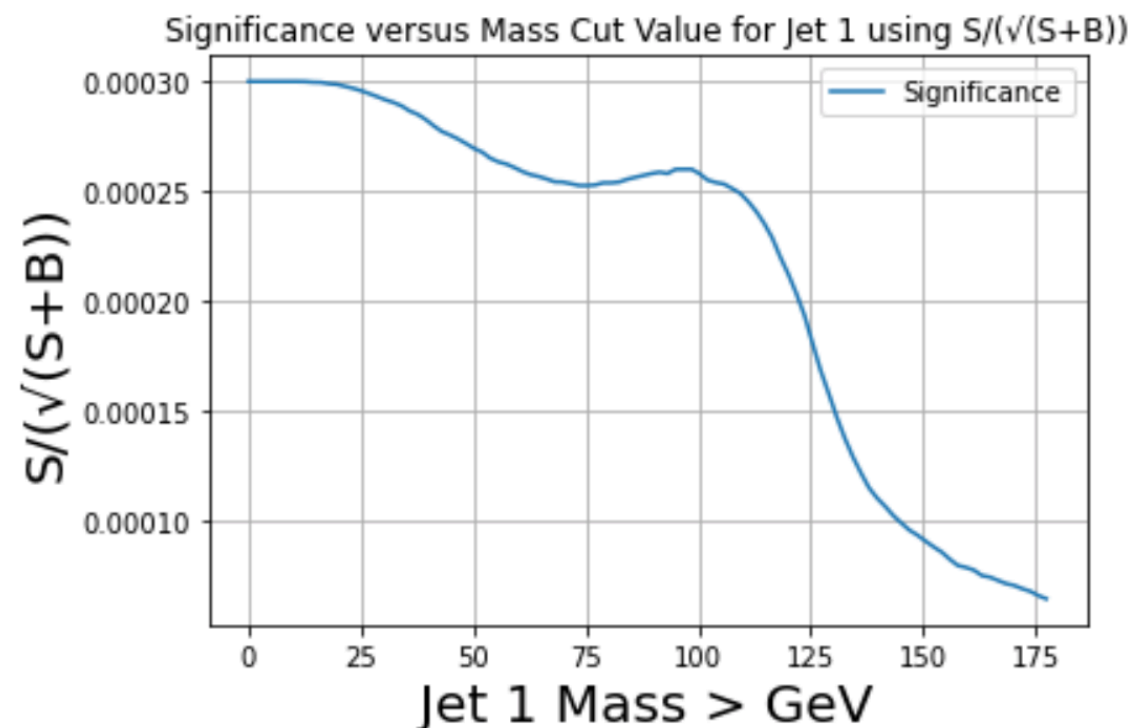
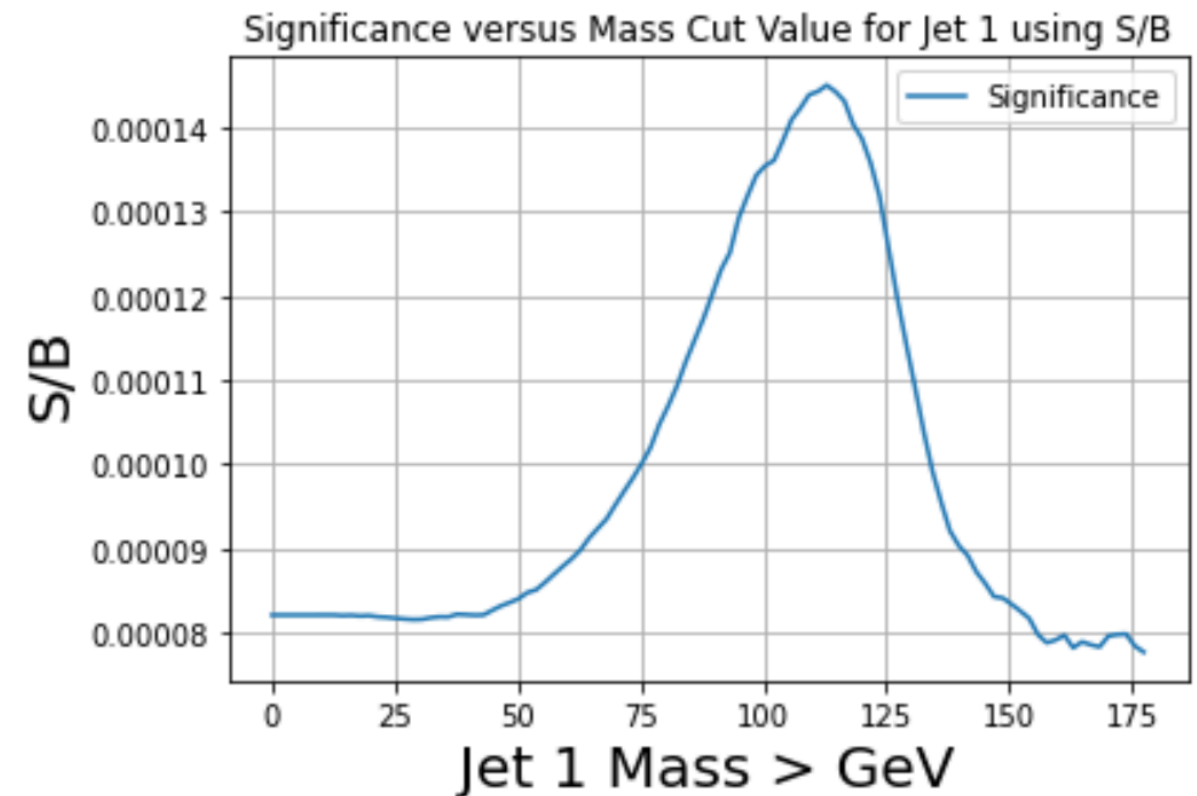
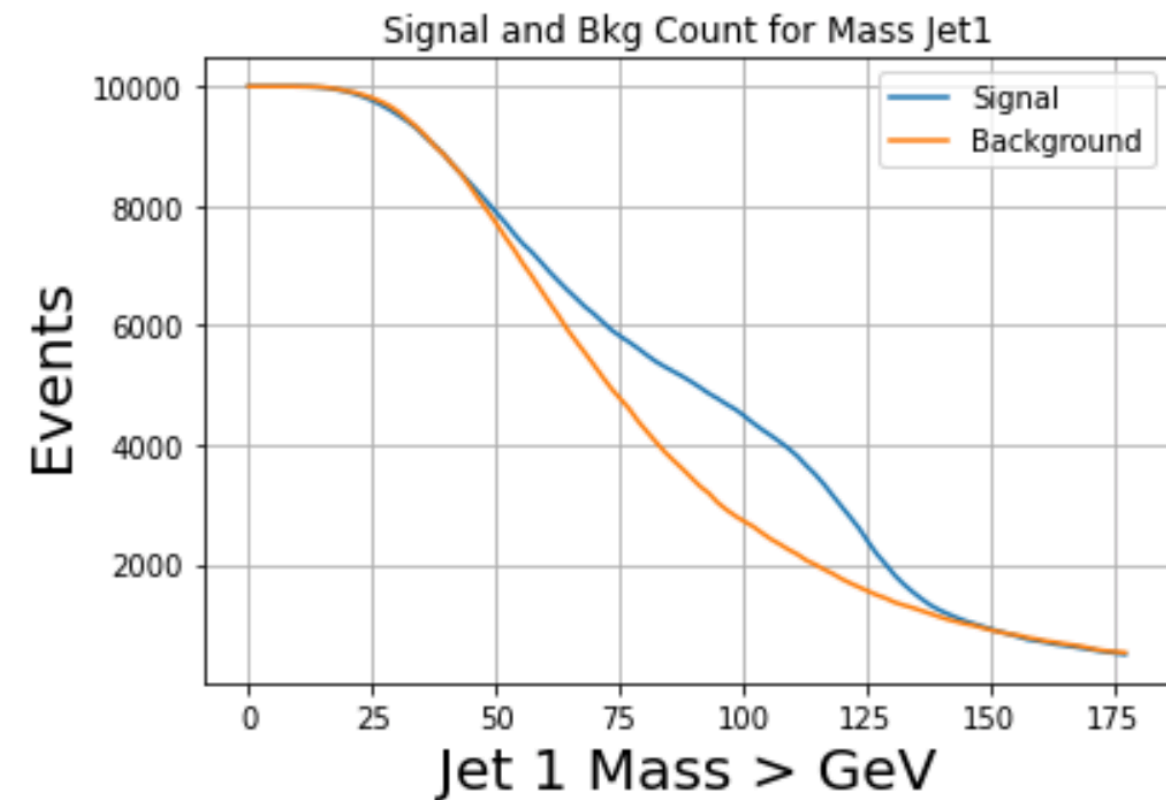


CMS PAS B2G-21-001



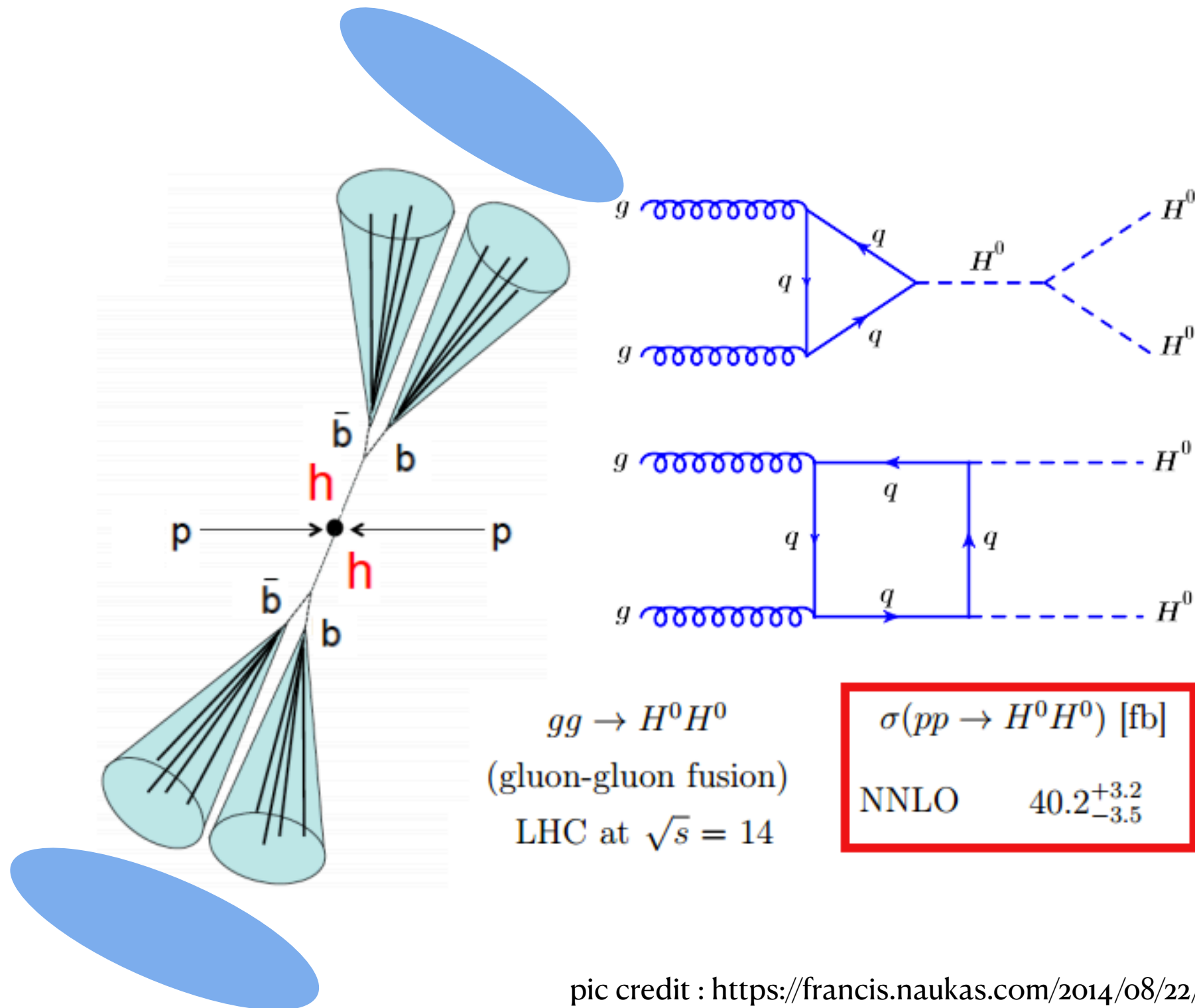
# Classify events and scan the signal significance

## Significance vs Mass



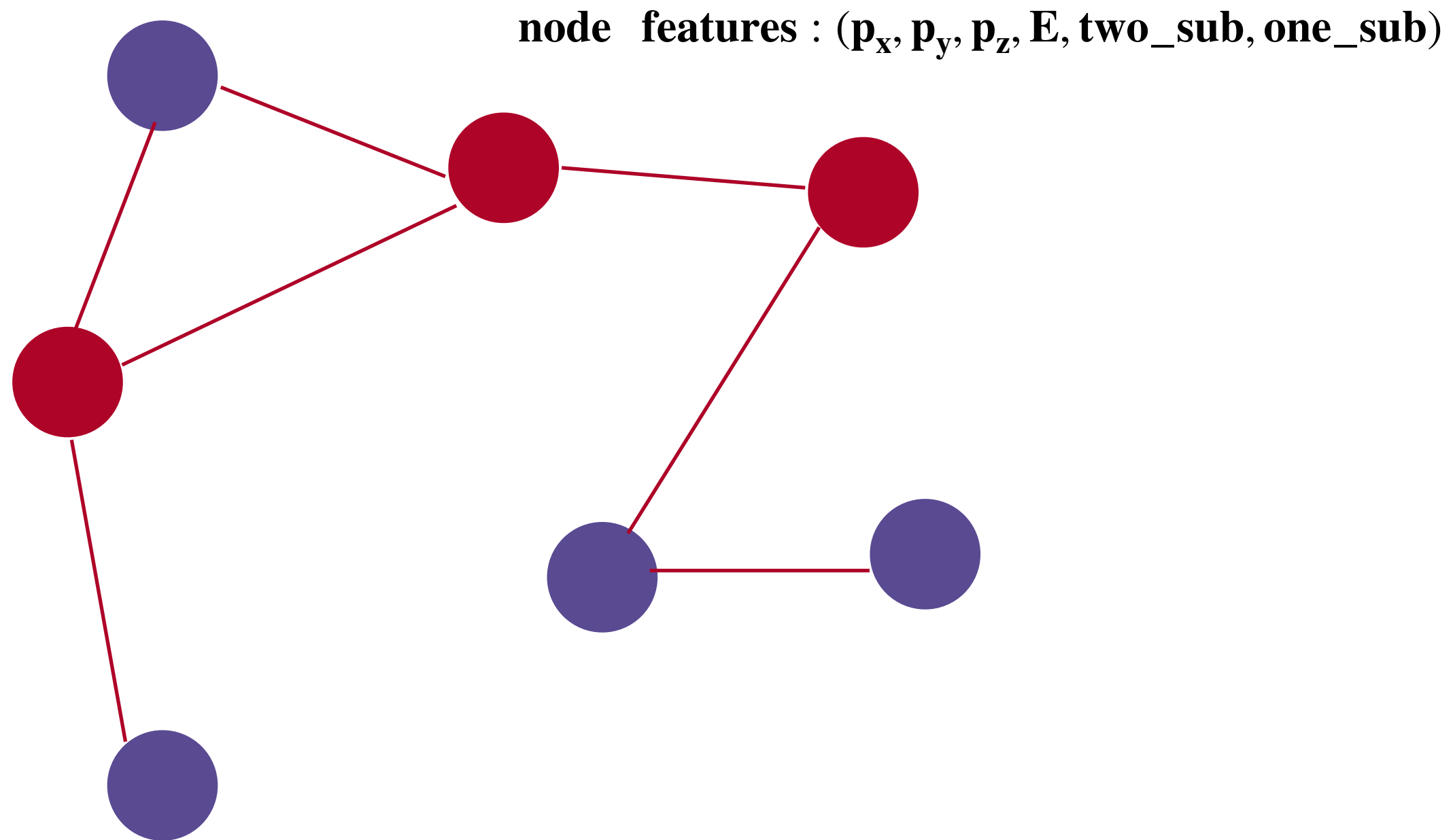
# Event classification using GNN

For boosted di-Higgs production we look for two Ak-08 jets with track subjects



pic credit : <https://francis.naukas.com/2014/08/22/el-campo-de-higgs/>

# Event as a graph



Locate all the  $R=0.8$  PF jets and  $0.4$  track jets in the  $\eta, \phi$  plane.  
Connect the  $k$ -NN neighbour through edges.  
For each nodes : assign 4-vector + two and one subjettiness observables.  
Use this graph representation for the events to be fed in GNN.



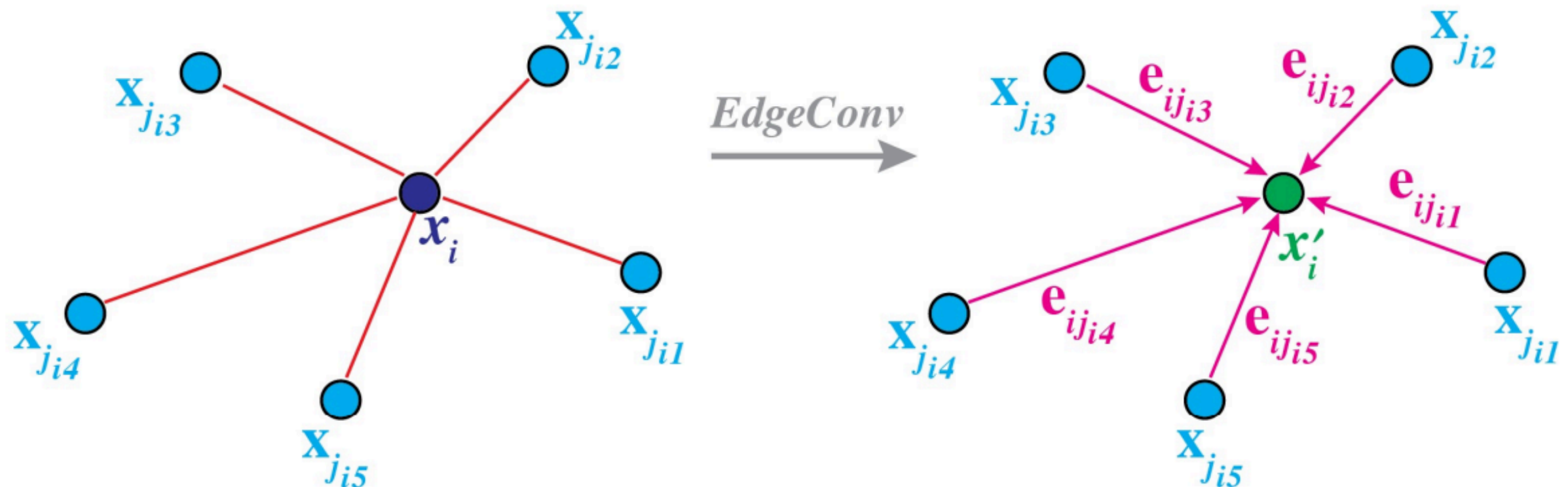
# The graph network

<https://arxiv.org/pdf/1801.07829.pdf>

In a graph, each node can “learn” about the state of neighboring node through message passing operation

$$(x')_i^{l+1} = \max_{j \in \mathcal{N}(i)} \Theta_x(x_j^l - x_i^l) + \Phi_x(x_i^l)$$

$$(e')_i^{l+1} = \text{mean}_{j \in \mathcal{N}(i)} \Theta_e(e_j^l - e_i^l) + \Phi_e(e_i^l)$$

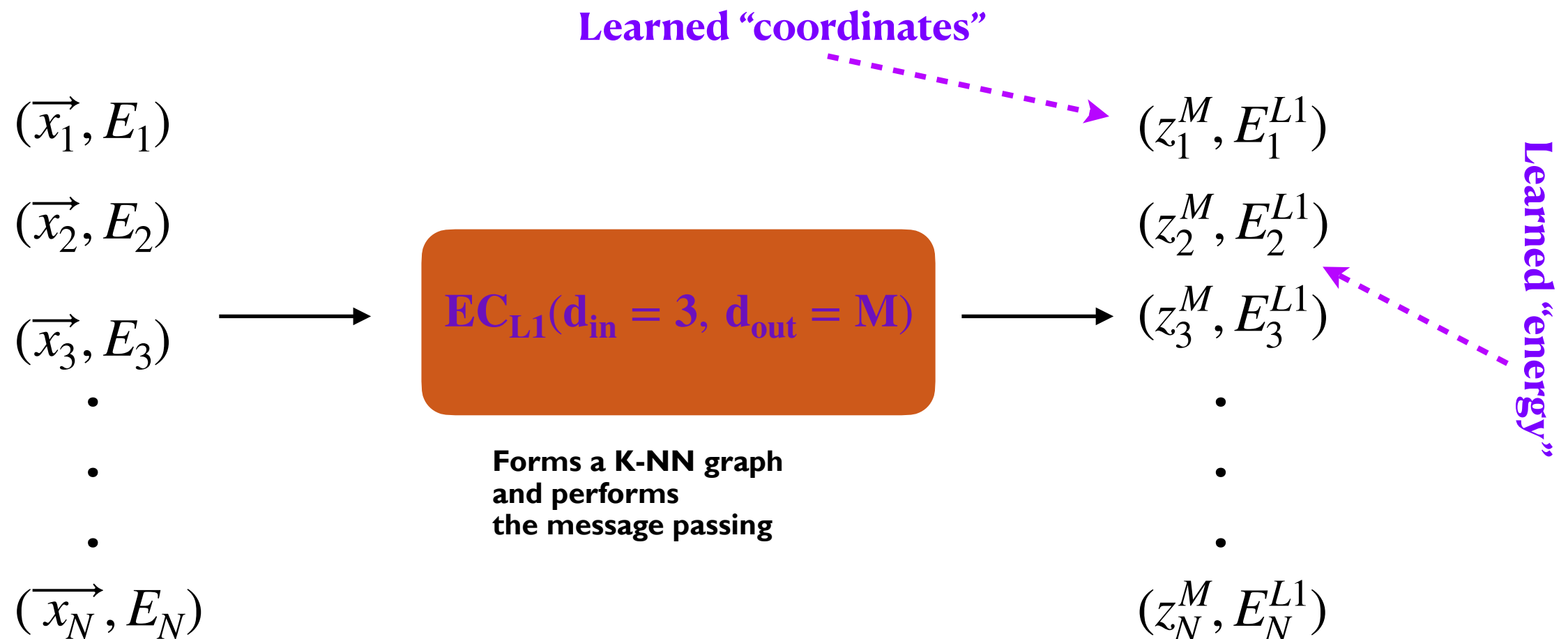


# The graph network

<https://arxiv.org/pdf/1801.07829.pdf>

$$(x')_i^{l+1} = \max_{j \in \mathcal{N}(i)} \Theta_x(x_j^l - x_i^l) + \Phi_x(x_i^l)$$

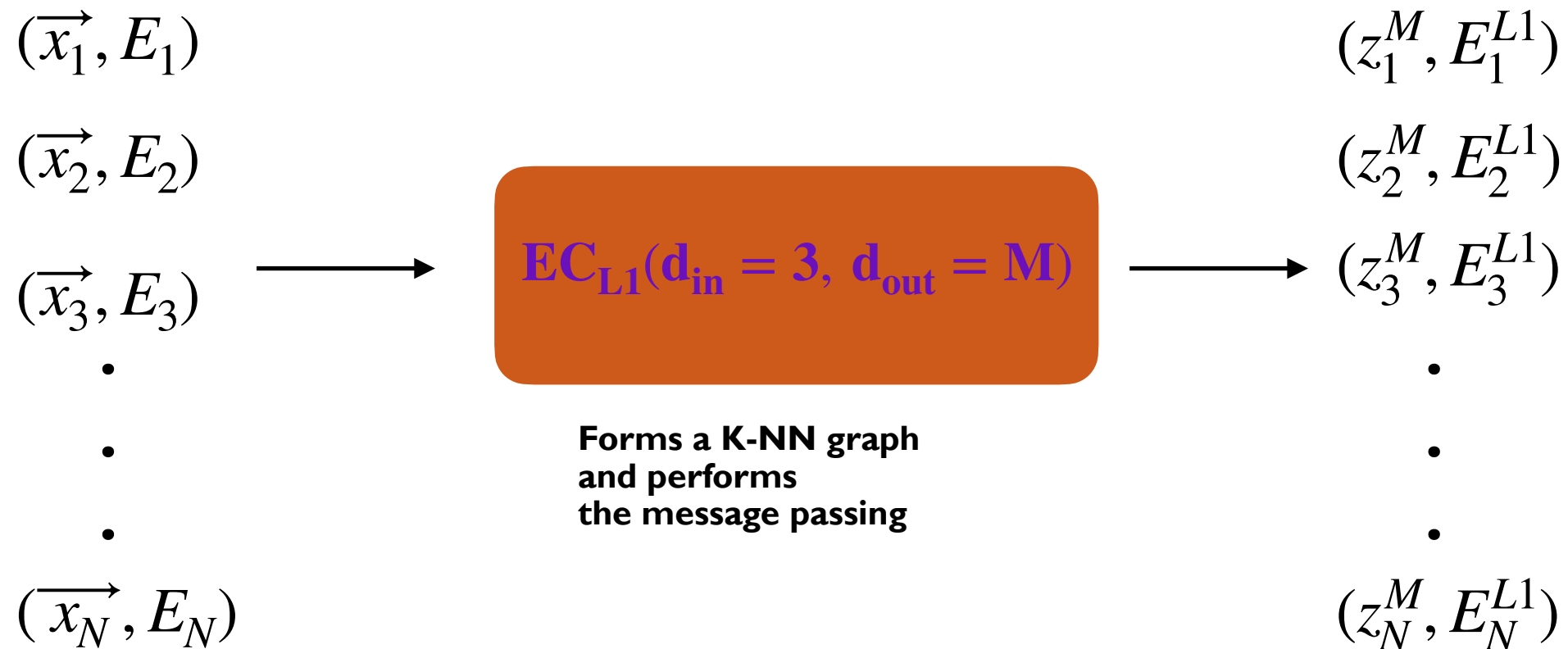
$$(e')_i^{l+1} = \text{mean}_{j \in \mathcal{N}(i)} \Theta_e(e_j^l - e_i^l) + \Phi_e(e_i^l)$$



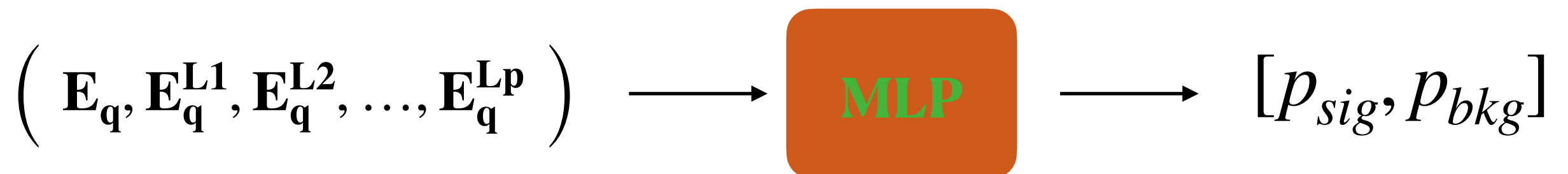


# The graph network

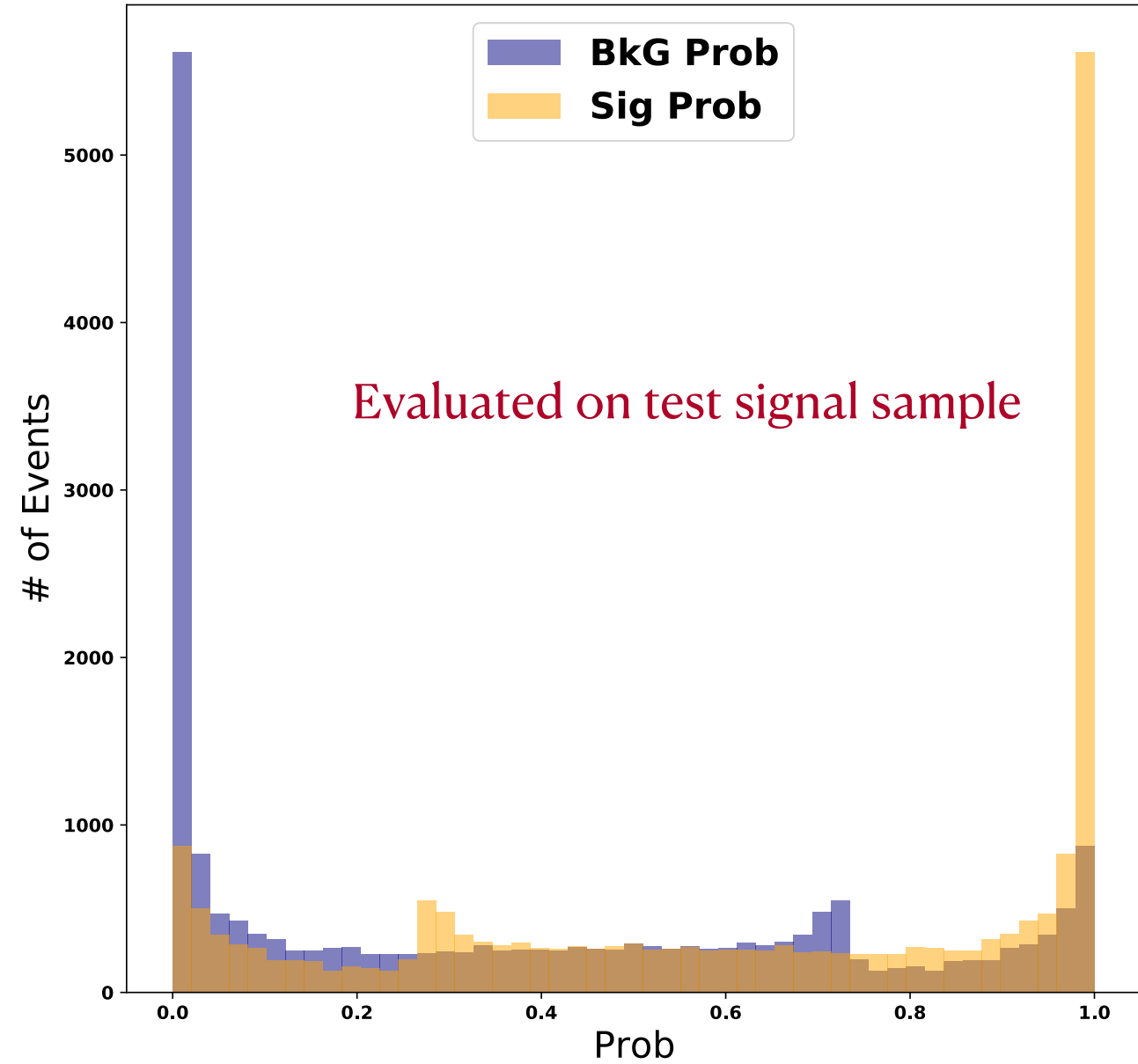
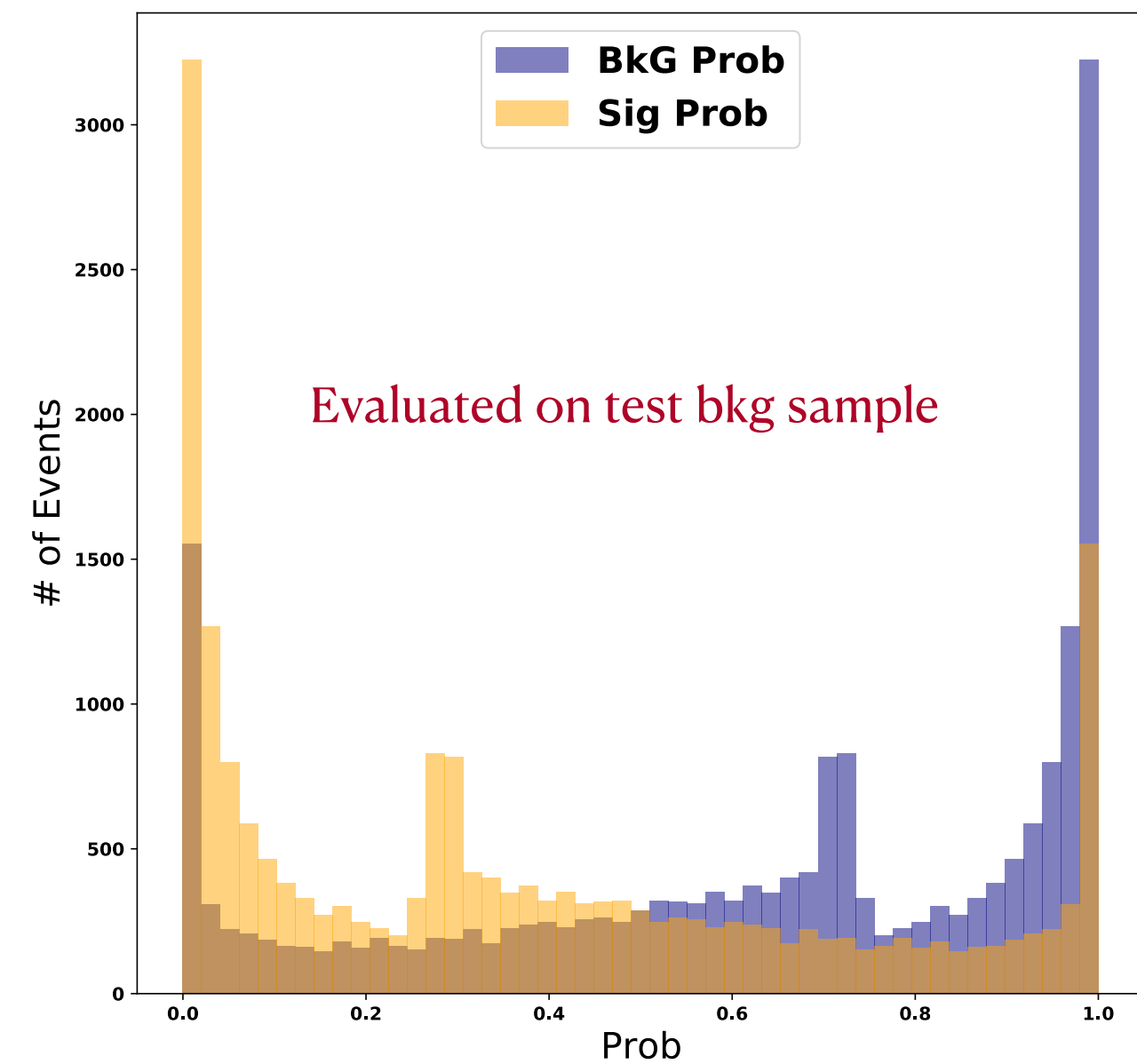
<https://arxiv.org/pdf/1801.07829.pdf>



After  $p$  message passing layers, the  $q$ -th node has following energy representation :



# Signal vs QCD bkg separation using GNN



# Discussion

- ☑ Ultra boosted Di-Higgs production is an interesting physics case study for FCC
- ☑ We are performing a feasibility study with a ML (GNN) based method to increase the signal sensitivity.
- ☑ Initially looking into event classification : a demerit of the method would be no trivial way to extract the calibration factors.
- ☑ Eventually, will use all the object level features in order to properly evaluate scale factor for the events.
  - Representing events as heterograph might appear handy in this context.
  - Will perform comparison of different NN models.
- ☑ Finally we wish to build an interpretable NN pipeline in order to relate individual physics observables which contribute towards the NN performance.

# Backup

# Classify events and scan the signal significance

heatmap for significance of invariant mass and pt jet1 using S/B

